DOCUMENT RESUME

ED 142 558 TM 005 772

AUTHOR Forsyth, Robert A.; Feldt, Leonard S.
TITLE On the Validity of the ITED as an Aid in Program Evaluation. Iowa Testing Programs Occasional Papers No. 3.
INSTITUTION Iowa Testing Programs, Iowa City.
REPORT NO ITP-OP-3
PUB DATE Jun 73
NOTE 19p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Achievement Tests; Curriculum Evaluation; *Educational Objectives; *Educational Programs; *Evaluation Criteria; Evaluation Methods; Measurement Techniques; Program Evaluation; Secondary Education; Standardized Tests; Test Interpretation; *Test Validity
IDENTIFIERS *Iowa Tests of Educational Development

ABSTRACT

A brief discussion of the concept of validity, a description of the nature and purposes of the Iowa Tests of Educational Development (ITED), and a rationale for the inclusion of ITED results as part of the overall program evaluation of a secondary school are presented. According to the authors, the ITED can validly be used as one of the data-gathering instruments for program evaluation if the evaluation instruments are judged on the basis of the behaviors that they require of students. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED142558

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IOWA TESTING PROGRAMS OCCASIONAL PAPERS

Number 3 - June 1973

On the Validity of the ITED as an Aid in Program Evaluation

Robert A. Forsyth and Leonard S. Feldt

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Robert A. Forsyth

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND THE ERIC SYSTEM CONTRACTORS."

TM005 772

ON THE VALIDITY OF THE ITED AS AN AID IN PROGRAM EVALUATION*

Robert A. Forsyth Leonard S. Feldt

Can the ITED be validly used as one of the data-gathering instruments in the evaluation of secondary school programs? The authors think they can. The acceptability of the ITED for this purpose depends, however, on a particular philosophy of evaluation. This philosophy holds that evaluation instruments must be judged on the basis of the behavior they require of students. The hallmark of adequate evaluation instruments is a "close fit" between the skills that students use on the tests and the skills that are the goals of the program. Similarity of test materials and local instructional materials is not a crucial consideration. Under this philosophy all aspects of a program, including the curriculum itself, is subject to evaluation.

Introduction

Until recent years the public rarely challenged the judgment of professional educators regarding the return on the investment in public education. Such challenges are becoming increasingly common. [See, for example, Lessinger (1970) and Dyer (1973).] In many communities interested citizens are asking, "Are we getting our money's worth from what we are spending for our schools?" Vague reassurances that expenditures are worthwhile are not being accepted in the absence of factual evidence. As a result, increasing attention is being given to the problems of program evaluation.

---

*Parts of this paper are taken from the Manual for Teachers, Counselors, and Examiners, Forms X-6 and Y-6, ITED and the Manual for Administrators and Testing Directors, ITED, Forms X-6 and Y-6, published by the Iowa Testing Programs, University of Iowa.

Even a cursory examination of the evaluation literature leads to the conclusion that a "good" evaluation is an extremely time-consuming undertaking. First, it is necessary to identify which of the many desirable objectives of instruction shall be emphasized in a particular evaluation study. It is also necessary to evaluate relevant input variables--the nature of the pupils entering the system, the funds available to support the program, the adequacy of buildings and equipment, etc. Next, it is necessary to select or develop measures of outcomes (both intended and unintended), to administer the various instruments, and to compile the relevant data. Finally, the data must be analyzed and interpreted. Each of these steps in the evaluation process requires many hours of reflection and effort.

The time required to develop measures of outcomes can be exceedingly great, if careful tryout and refinement of materials are undertaken. It is understandable, therefore, that administrators and evaluation committees generally prefer to adopt existing instruments rather than develop original tests, inventories, questionnaires, and rating scales. Because of their wide use in Iowa high schools, the Iowa Tests of Educational Development are an obvious possibility for evaluating some of the cognitive outcomes of secondary programs. Can they be validly used for this purpose? Within the limitations noted below, we believe they can.

In making a case for the ITED, we would first note that specifying the limits of any program may not be as easy as it seems. One could say, probably without controversy, that an educational program consists of the activities a school faculty employs to accomplish a given set of objectives. But within this definition, differences in philosophy may exist unnoticed. Some may view a program broadly and feel it includes practically every student experience--structured or unstructured--that contributes to the objectives under study. Others may conceive of a program more narrowly and view it in terms of those activities specifically planned to produce the desired outcomes. For example, the science program could be defined to include any experience, in school or out, which adds to student understanding of the nature of the universe and the work of scientists. Alternatively, the program could be defined to include only those elements and activities which occur in science classes and are explicitly controlled by the teacher. On another dimension, one can be concerned with only that portion of the program required of all students or with every aspect--the remedial levels, the common core of experiences to which all students are exposed, and the advanced or specialized activities intended for relatively few. Thus, the scope of the program to be evaluated and the limits of the school's responsibility are issues on which disagreement may exist. As we will try to show, the definition of a program that

one adopts has implications for the usefulness of the ITED--or any other measure--as an aid in evaluation.

To assist you in deciding whether or not there is a place for the ITED in your program evaluation efforts, we present (1) a brief discussion of the concept of validity as it is currently viewed by educational measurement specialists, (2) a brief description of the nature and purpose of the ITED, and finally, (3) a rationale for the inclusion of the ITED results as a part of the overall program evaluation.

1. Concept of Validity

One occasionally hears a teacher or administrator state categorically, "The ITED just aren't valid for our school." The degree of truth in this statement, as it stands, is impossible to determine. A test (or test battery) is probably never totally valid or invalid. As Gronlund (1971, p. 77) states:

> Validity pertains to the *results* of a test, or evaluation instrument, and *not* to the instrument itself. We sometimes speak of the validity of a test for the sake of convenience, but it is more appropriate to speak of the validity of the interpretation to be made from the results.

Gronlund also indicates two additional cautions concerning the concept of validity (1971, p. 77):

> Validity is a *matter of degree*. It does not exist on an all-or-none basis. Consequently, we should avoid thinking of evaluation results as valid or invalid. Validity is best considered in terms of categories that specify degree, such as high validity, moderate validity, and low validity.

> Validity is always *specific to some particular use*. It should never be considered a general quality. For example, the results of an arithmetic test may have a high degree of validity for indicating computational skill, a low degree of validity for indicating arithmetic reasoning, a moderate degree of validity for predicting success in future mathematics courses, and no validity for predicting success in art or music. Thus, when appraising or describing validity, it is necessary to consider the use to be made of the results. Evaluation results are never just valid; they have a different degree of validity for each particular use to which they are put.

Cronbach expresses a similar view in a very few words (1971, p. 443): "Validation examines the soundness of all interpretations of a test . . ." There are many implications of this simple statement. How many interpretations of the test results can be made? If we are considering a teacher-made test given at the end of a unit of instruction for the purpose of assigning grades on that unit, perhaps the number of interpretations is limited. Or, if a diagnostic test is given for the purpose of identifying necessary areas of work for students, a single purpose is implied. However, when a standardized test such as the ITED is given, the number of possible interpretations of the results is potentially much greater. The ITED, like most standardized tests, are designed to serve a variety of purposes. For example, the tests are useful for identifying general strengths and weaknesses of individual students, in identifying over- and under-achievers, in program evaluation, and in educational guidance. Furthermore, it is possible to identify at least six different groups of people

who might be interested in using the scores: (1) students; (2) parents; (3) teachers; (4) counselors; (5) school board members; and (6) administrators. The science scores, for example, may be utilized by the counselor to help predict success in future science courses at the college level (or high school), and the same scores may be used by the administrator to help in the evaluation of the science program. Obviously, the degree of validity of the scores for each purpose needs to be determined before one can have confidence in an interpretation.

Although test publishers frequently supply a large amount of "validity" evidence, it is the responsibility of any school system utilizing any test to concern itself with validity of the particular uses actually being made locally. Cronbach (1970, p. 36) emphasizes this idea when he states, "*Validation is the task of the test interpreter.* Others (i.e., publishers or measurement specialists) can do no more than offer him material to incorporate into his thinking." The primary purpose of this paper is to supply some of this "thinking material" to the administrators and teachers in Iowa.

2. Nature and Purpose of ITED

A detailed description of ITED content, including the classification of each item according to the objective being measured, is given in the Manual for Teachers, Counselors,

and Examiners for Forms X-6 and Y-6. Perhaps the best way to understand the nature of the battery, however, is to actually take the tests. This point cannot be emphasized too greatly. It has been our experience that in many instances where teachers have stated that the tests were not valid they had not even examined the test items. It is our belief that each subtest should be thoroughly examined by the teachers concerned with the subject area being measured. If such an examination leads to the conclusion that, in general, the items aren't measuring important outcomes, then the results cannot be valid for any purpose. However, we feel that such an examination will support our contention that important objectives are being measured. A thorough examination of the tests should serve as a first step to better utilization of test results.

The ITED attempt to measure abilities that are important in adult life and constitute the foundation for continued learning. These skills include the ability to recognize the essentials of good writing, to resolve quantitative problems, to weigh discussions of social issues critically, to recognize sound methods of scientific inquiry, to perceive the subtle meanings and moods of literary materials, and to use sources of information.

The ITED are achievement tests in the broadest sense. They require the student to use his knowledge and skills in

analyzing materials that he probably has not encountered previously. Thus, the tests are designed to measure how well the student can apply his education in new settings. Only in this sense are the tests concerned with the specific knowledge and skills that constitute the immediate objectives of individual high school courses.

The ITED battery is intended to provide measures of educational achievement that are appropriate for the very large majority of high school students, regardless of the particular curriculum they are following. Clearly, each student has certain unique objectives, needs, and interests with which his teachers are concerned. But students also have many needs in common. Individualization in education generally concerns methods and materials, not differentiation in long-range educational objectives. The authors of the ITED have attempted to look beyond the immediate means by which various goals might be attained and to concentrate upon the intellectual behavior represented by the goals themselves. Thus, they believe the tests to be appropriate in an era which emphasizes diversity of educational programs.

The ITED are not intended to serve the functions of final examinations. This is an important point for a school faculty to appreciate. There is a real need for measurement of the immediate outcomes of various high school courses, but as the diversity of instructional methods and materials increases--from school to school and from pupil to pupil

within a school--standardized tests become less and less appropriate for this purpose. Such tests serve a more valuable function, the authors believe, when they concentrate on the goals toward which various methods and materials converge.

3. Using ITED Results as a Part of Program Evaluation: A Rationale

Educational administrators generally look upon the school testing program as one of the important administrative tools in the evaluation of the local educational program. However, many teachers feel that validity for this purpose is extremely limited. They take the position that if tests are to be used for the purpose of evaluating the educational program they must conform very closely to the content of the local curriculum. Instruments administered to any student must be based on those courses he personally has taken. Moreover, the resultant data can be legitimately compared only with that accumulated in schools following a very similar curriculum and drawing upon a similar student population.

A fundamental premise of this philosophy is the belief that evaluation should be concerned only with how well the locally adopted goals have been achieved in each subject area. The test results, it is argued, should not

be used to challenge the legitimacy of these objectives or the methods being followed to achieve them. In fact, according to this point of view, the most adequate instruments should conform to these goals and methods in all important respects. As one might infer, the tests most favored under this philosophy tend to emphasize the most immediate goals of instruction and to reflect the local choice of methodology and course content. Since standardized tests would rarely satisfy these demands, such tests are generally seen to have limited worth for program evaluation.

Teachers holding this philosophy want achievement tests to include a generous sampling of the particular content that constitutes much of their day-to-day concern. They may be critical of tests that do not contain exercises patterned after the local curriculum materials. For example, teachers convinced of the value of the linguistic approach to the teaching of language arts may demand tests containing exercises specific to that approach. They may not be content to accept a test that ignores the instructional approach and is concerned solely with the student's ability to use language effectively. In such instances, teachers often voice their discontent with standardized tests by stating, "These tests are not measuring what we're teaching."

There is, however, another philosophical position that is held by many educators. Proponents of this position would argue that evaluation procedures should assess progress toward all objectives that are viewed as important by responsible educators and laymen. Cronbach, for example, feels that there are times when the evaluation procedures should assess the attainment of outcomes beyond those which have been established for a given course or program. He writes (1963, p. 680):

> In course evaluation, we need not be much concerned about making measuring instruments fit the curriculum. However startling this declaration may seem, and however contrary to the principles of evaluation for other purposes, this must be our position if we want to know what changes a course produces in the pupil. *An ideal evaluation might include measures of all the types of proficiency that might reasonably be desired in the area in question, not just the selected outcomes to which this curriculum directs substantial attention.* [Italics added] If you wish only to know how well a curriculum is achieving its objectives, you fit the test to the curriculum; but if you wish to know how well the curriculum is serving the national interest, you measure all outcomes that might be worth striving for. One of the new mathematics courses might disavow any attempt to teach numerical trigonometry, and indeed, might discard nearly all computational work. It is still perfectly reasonable to ask how well graduates of the course can compute and can solve right triangles. Even if the course developers went so far as to contend that computational skill is no proper objective of secondary instruction, they will encounter eductors and laymen who do not share their view. If it can be shown that students who come through the new course are fairly proficient in computation despite

> the lack of direct teaching, the doubters will be reassured. If not, the evidence makes clear how much is being sacrificed.

More recently, Cronbach has stated (1971, p. 460):

> The recommendation that the evaluation battery be comprehensive seems to run counter to the concept that an educational test should measure what has been taught. And students think a test "unfair" when it asks about topics not covered in the course. One can agree that it is unjust to let the fate of an individual be determined by a test for which, through no fault of his own, he is ill-prepared. But this only illustrates once more how a test valid for one decision can be invalid for another. *Though it is unfair to judge the quality of a teacher's work by a test that does not fit the course of study he was directed to follow, that test may be a fair basis for judging the curriculum.* [Italics added] If teacher plus course-of-study have left the pupil ignorant of contemporary literature, this is a significant fact about the adequacy of his education.
>
> Sometimes a test can "fit the curriculum" entirely too well. If the key to a test in literary comprehension gives credit only for an "authorized" interpretation that the teacher has handed down to the students, it tells nothing about their ability to interpret literature . . . The universe pertinent in summative evaluation is the universe of tasks graduates are expected to perform. To be sure, a curriculum developer who has a restricted objective can use a restricted test to determine how well he achieved his end. But if other educators considering adoption of the course desire outcomes that go beyond his aims, they will find his studies inadequate.
>
> The primary reason for examining the whole range of outcomes that interest responsible educators is to maximize the soundness of evaluative conclusions. The effect of such measurement upon teachers and students is a further advantage. Teachers who honestly intend to cover a whole long list of objectives find that class time is insufficient to pursue them all with equal zeal. They are most likely to sacrifice those objectives for which no

> evaluation data will be collected. Similarly, the student, in deciding what to study and how, is strongly influenced by his perception of what "counts." Any broadening of the evaluation procedures is therefore likely to have a healthy educational effect.

The use of the ITED for program evaluation can be more strongly defended if the second philosophy is accepted as reasonable. According to this view, if groups of educators and/or laymen feel that important proficiencies are being measured by the ITED, then the results are valid regardless of how closely the actual items can be identified with specific lessons and activities in the local curriculum. We believe that in most communities in Iowa there would be no conflict between the local curriculum objectives and the objectives being measured by the test.*

Even for school systems where the staff holds to the first philosophical position, the tests usually will be

---

*If the tests are relevant, the question may be raised, "Why can't instruments measuring the same objectives as the ITED be constructed locally?" Such instruments could be built, of course. However, good tests are difficult and expensive to build. And if the items on the ITED do measure important objectives, then the use of this standardized instrument offers schools opportunities for both a norm-referenced and a criterion-referenced interpretation of the scores. That is, not only can the school interpret the results internally in an absolute sense, but normative comparisons can also be made.

found to have a high degree of validity. It has been our experience that in the majority of instances where teachers have examined the subtests of the ITED related to their teaching area they have concluded that the tests are measuring important objectives--objectives they want their students to attain. Much of the criticism of using standardized tests in program evaluation has been related to the use of such tests as the only evaluation data. Thus, there is good reason to believe that when teachers say, "The tests are not measuring what we're teaching," they have not stated their feelings quite accurately. A more appropriate statement would be, "The tests aren't measuring all that we teach," or, "The tests emphasize many long-range objectives, and we would like more attention given to the specific objectives of our classes."

Certainly, it must be realized that any test such as the ITED cannot measure all the worthwhile outcomes of a given educational program. If the test results constitute the only data to evaluate a program, then for that purpose they have low validity. For example, no multiple-choice test can measure the student's ability to write a well-organized essay on a given topic. Thus, to use the results of the ITED subtest in effectiveness of expression as the only evidence of program success would be ridiculous. However, if the results are utilized as that part of the

evaluation data bearing on the specific objectives being measured by the ITED, they have a high degree of validity.

Some Concluding Remarks

Throughout this paper we have consistently tried to refer to the ITED results as a part of the evaluation data. No single test tells everything about a school system. Different kinds of tests yield different kinds of insight, and the importance of one does not diminish the importance of another. Many facts essential to valid program evaluation are poorly revealed by all instruments presently available. No measures reliably assess attitudes toward and commitment to social change, for example, or ability to work with others for political action. The tendency to evaluate a complex enterprise solely on the basis of a few facts is as foolish in education as it is in government or public health.

This paper has suggested ways in which the ITED might be justified as part of the program evaluation effort. The discussion has been very general and has not focused on any specific area. It should be obvious that for the evaluation of a number of important programs, such as those involving many specific vocational skills, the ITED has no validity. Nor have we discussed the extremely important question, "How does one use the results for program evaluation?" Several suggestions related to the "how to" aspects are

given in the Manual for Administrators and Testing Directors, ITED, Forms X-6 and Y-6. These suggestions are of both a norm-referenced and a criterion-referenced nature. This manual also contains a discussion of some of the cautions that must be observed when utilizing ITED data as a part of any program evaluation.

Finally, we would like to repeat an earlier idea. If the ITED are being given in your school, and if you are using them as a part of your program evaluation efforts, your decision should be validated. Teachers and others (both educators and the lay public) should examine the tests. They should agree that the objectives being measured by the ITED are important. If this is done and if they agree on the validity (to some degree) of the ITED results as a part of the program evaluation effort, then the entire ITED program has a better chance for being successful.

References

Cronbach, L. J. Evaluation for Course Improvement. Teachers College Record, 1963, 64, 672-683.

Cronbach, L. J. Validation of Educational Measures. Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton: Educational Testing Service, 1970.

Cronbach, L. J. Test Validation. In R. Thorndike (ed.), Educational Measurement (2nd ed.). Washington: American Council on Education, 1971. pp. 443-507.

Dyer, H. S. How to Achieve Accountability in the Public Schools. Bloomington, Indiana: Phi Delta Kappa Educational Foundation, 1973.

Gronlund, N. E. Measurement and Evaluation in Teaching. New York: The Macmillan Company, 1971.

Lessinger, L. M. Every Kid A Winner: Accountability in Education. Palo Alto: Science Research Associates, 1970.